



Munich Personal RePEc Archive

The Discursive Dilemma and Probabilistic Judgement Aggregation

Pivato, Marcus

Department of Mathematics, Trent University

23. April 2008

Online at <http://mpra.ub.uni-muenchen.de/8412/>

MPRA Paper No. 8412, posted 23. April 2008 / 20:56

The Discursive Dilemma and Probabilistic Judgement Aggregation

Marcus Pivato

Department of Mathematics, Trent University,
Peterborough, Ontario, Canada
marcuspivato@trentu.ca

April 23, 2008

Abstract

Let \mathcal{S} be a set of logically related propositions, and suppose a jury must decide the truth/falsehood of each member of \mathcal{S} . A ‘judgement aggregation rule’ (JAR) is a rule for combining the truth valuations on \mathcal{S} from each juror into a collective truth valuation on \mathcal{S} . Recent work has shown that there is no reasonable JAR which always yields a logically consistent collective truth valuation; this is referred to as the ‘Doctrinal Paradox’ or the ‘Discursive Dilemma’.

In this paper we will consider JARs which aggregate the subjective *probability* estimates of the jurors (rather than Boolean truth valuations) to produce a collective probability estimate for each proposition in \mathcal{S} . We find that to properly aggregate these probability estimates, the JAR must also utilize information about the private information from which each juror generates her own probability estimate.

Suppose there are three propositions, A and B , and “ $A \Rightarrow B$ ”, and a jury with three jurors $\mathcal{J} = \{1, 2, 3\}$, which must decide the truth or falsehood of these propositions by ‘aggregating’ the judgements of the individual jurors.¹ Suppose the jurors have the profile of truth-valuations shown in Table 1(A). Each juror has a logically consistent truth-valuation, but the collective truth-valuation generated by majority vote is logically inconsistent. This is called the *Doctrinal Paradox* by Kornhauser and Sager (1986, 1993, 2004), because it can lead to logical inconsistencies in legal doctrine. This paradox is not merely an artifact of majority vote. List and Pettit (2002) have proved an ‘impossibility theorem’ which states (roughly) that there is no anonymous, neutral, and ‘systematic’ procedure which will aggregate any profile of juror truth-valuations into a logically consistent collective

¹Following the social choice literature, we will describe this problem as *judgement aggregation*. In the artificial intelligence literature, the same problem is studied as *belief merging*; see e.g. Cholvy (1998), Konieczny and Pino Pérez (2005), and Pigozzi (2006).

j	A	$A \Rightarrow B$	B
1	T	T	T
2	T	F	F
3	F	T	F
\mathcal{J}	T	T	F

(A)

j	A	$A \Rightarrow B$	B
1	$\mu_1[\mathbf{A}]$	$\mu_1[\mathbf{A}^c \cup \mathbf{B}]$	$\mu_1[\mathbf{B}]$
2	$\mu_2[\mathbf{A}]$	$\mu_2[\mathbf{A}^c \cup \mathbf{B}]$	$\mu_2[\mathbf{B}]$
3	$\mu_3[\mathbf{A}]$	$\mu_3[\mathbf{A}^c \cup \mathbf{B}]$	$\mu_3[\mathbf{B}]$
\mathcal{J}	$\mu_{\mathcal{J}}[\mathbf{A}]$	$\mu_{\mathcal{J}}[\mathbf{A}^c \cup \mathbf{B}]$	$\mu_{\mathcal{J}}[\mathbf{B}]$

(B)

j	$\mu_j[\mathbf{A}]$	$\mu_j[\mathbf{A}^c \cup \mathbf{B}]$	$\mu_j[\mathbf{B}]$
1	α_1	γ_1	β_1
2	α_2	γ_2	β_2
3	α_3	γ_3	β_3
\mathcal{J}	$\frac{\alpha_1 + \alpha_2 + \alpha_3}{3}$	$\frac{\gamma_1 + \gamma_2 + \gamma_3}{3}$	$\frac{\beta_1 + \beta_2 + \beta_3}{3}$

(C)

Table 1: **(A)** A profile of truth-valuations yielding a Discursive Dilemma. For example, suppose A = “Atmospheric CO₂ will rise to 710 ppm by 2100 AD”, B = “Average global temperature will rise by 5.5° C by 2100 AD”, while “ $A \Rightarrow B$ ” = “If atmospheric CO₂ rises to 710 ppm, then average global temperature will rise by 5.5° C”. **(B)** A probabilistic version of Table (A). **(C)** The Linear Opinion Pool for Table (B), with $w_1 = w_2 = w_3 = 1/3$;

truth-valuation.² To emphasize its ramifications to collective discourse in general (not just legal doctrine), List and Pettit (2002) call this the *Discursive Dilemma*.

This model of judgement aggregation uses Boolean logic, which is only appropriate if each juror holds her truth-valuation with absolute certainty. However, these truth-valuations contradict one other, so at least one juror must be wrong —hence at least one juror cannot really be ‘certain’ in her beliefs. Indeed, outside of mathematics, there is really no sphere of knowledge where people can make assertions with absolute certainty. Normally each person has some degree of ‘confidence’ in a belief—a subjective probability which is strictly between zero and one—based on the quantity and quality of evidence available to her.

Thus, Pauly and van Hees (2006) and van Hees (2007) extended List and Pettit’s Discursive Dilemma theorem to \mathcal{T} -valued logics, where $\mathcal{T} = \{0, 1, \dots, T\}$ is a finite set of truth values: ‘ T ’ represents ‘true’, ‘0’ is ‘false’, and the intermediate values are various ‘degrees of acceptance’ or ‘degrees of truthfulness’. Similarly Gärdenfors (2006) proved the Discursive Dilemma when each juror is allowed to ‘withhold judgement’ on one or more propositions (this can be seen as a 3-valued logic, with $\mathcal{T} = \{F, ?, T\}$). Finally Dietrich (2007) extended the Discursive Dilemma to a variety of generalized logics, including modal logics containing a modal operator \Box (where $\Box A$ means ‘it is quite probable that A ’).

However, it is difficult to operationalize the meaning of \mathcal{T} -valued logics (what exactly are ‘degrees of acceptance’?) or the modal operator ‘ \Box ’ (how probable is ‘quite probable’?). Thus, we suggest that a more realistic model of judgement aggregation should go beyond pure logic (Boolean, multivalued, modal, or otherwise), and instead aggregate the subjective *probabilities* which the jurors assign to propositions.

Formally, let $(\mathbf{X}, \mathcal{B}, \mu)$ be a probability space^{3¶}, and let $\mathbf{A}, \mathbf{B} \subset \mathbf{X}$ be two events[¶]. The ‘true’ state of the universe is some unknown point $x \in \mathbf{X}$. The proposition A is the assertion

²Dietrich and List (2007) later replaced ‘systematicity’ with a weaker requirement of ‘independence’.

³To preserve continuity, probability theory is reviewed in an appendix. Terminology defined in the appendix is marked in the text with the symbol ‘¶’.

“ $x \in \mathbf{A}$ ”; the proposition B is the assertion “ $x \in \mathbf{B}$ ”; the proposition “ $A \Rightarrow B$ ” corresponds to the assertion “ $x \in \mathbf{A}^c \cup \mathbf{B}$ ”. For each $j \in \mathcal{J}$, let $\mathbf{K}_j \subset \mathbf{X}$ be an event representing j ’s ‘knowledge’. In other words, j knows (with certainty) that $x \in \mathbf{K}_j$ —but that is all she knows. Note that j ’s knowledgeability is inversely proportional to the size of \mathbf{K}_j : if $\mathbf{K}_j = \mathbf{X}$, then she is totally ignorant, whereas if $\mathbf{K}_j = \{x\}$, then she is omniscient. Thus, j ’s probability estimate of proposition A is the *conditional probability*[¶] $\mu_j[\mathbf{A}] := \mu[\mathbf{A}|\mathbf{K}_j]$. Likewise, we define $\mu_j[\mathbf{B}] := \mu[\mathbf{B}|\mathbf{K}_j]$, etc. The jury’s collective probability measure $\mu_{\mathcal{J}}$ is somehow determined by the conditional measures $\{\mu_j\}_{j \in \mathcal{J}}$ of the individual jurors; see e.g. Table 1(B). We now have two questions: **(Q1)** What is the best method to generate $\mu_{\mathcal{J}}$? **(Q2)** How should we interpret $\mu_{\mathcal{J}}$?

Sections 1-4 of this paper deal with **(Q1)**. §1 reviews the well-established statistical theory of ‘opinion pooling’ and its inadequacies; §2 contrasts this with an ideal of judgement aggregation through ‘full disclosure’ of private knowledge. Such ‘full disclosure’ is probably impossible in practice, so we next consider how much we can achieve without it. §3 reviews the theory of consensus via ‘common knowledge’ developed by Aumann (1976) and others. §4 presents a method to aggregate probability estimates using information about the degree of ‘independence’ between the knowledge of different jurors.

We then turn to **(Q2)**. §5 discusses the problem of ‘booleanizing’ a probabilistic judgement to get a Boolean truth valuation, and concludes that it is generally impossible and usually unnecessary anyways. §6 considers the implications of ‘unbooleanizability’ for the debate between ‘consequentialist’ and ‘deontological’ ethics.

1 Statistical opinion pooling

Let $\mathcal{P}(\mathbf{X})$ be the space of all probability measures[¶] on $(\mathbf{X}, \mathcal{B})$. If \mathcal{J} is a set of jurors, then a *statistical opinion pooling rule* (SOPR) is a function $\Phi : \mathcal{P}(\mathbf{X})^{\mathcal{J}} \rightarrow \mathcal{P}(\mathbf{X})$. If $\mu_j \in \mathcal{P}(\mathbf{X})$ is the subjective probability distribution of juror j (for all $j \in \mathcal{J}$), then $\mu_{\mathcal{J}} := \Phi[(\mu_j)_{j \in \mathcal{J}}] \in \mathcal{P}(\mathbf{X})$ is a probability distribution representing the ‘aggregated judgement’ of the jury. This problem was considered by Savage (1954, §10.2), and since then has generated an extensive literature; the survey article by Genest and Zidek (1986) lists 92 key papers in its annotated bibliography.

A popular SOPR is the *linear opinion pool* (LOP), defined by $\Phi[(\mu_j)_{j \in \mathcal{J}}] := \sum_{j \in \mathcal{J}} w_j \mu_j$, where $\{w_j\}_{j \in \mathcal{J}}$ are nonnegative ‘weights’ with $\sum_{j \in \mathcal{J}} w_j = 1$. Table 1(C) illustrates the LOP with $w_1 = w_2 = w_3 = 1/3$. Figure 2(A) is a Venn diagram labelled with the probabilities of various events. Table 3 shows Table 1(C) filled with the corresponding conditional probabilities. Figure 2(B) is a ‘density plot’ of the averaged measure $(\mu_1 + \mu_2 + \mu_3)/3$.

The LOP was originally suggested by de Finetti (1954) and Stone (1961); later, Lehrer and Wagner (1981) proposed it as a general framework for ‘rational consensus’ in epistemology, philosophy of science, semantics, ethics, and social choice. The LOP has several appealing characterizations (Genest and Zidek, 1986, §3). For example:

[i] Suppose Φ is an SOPR defined by a function $F : [0, 1]^{\mathcal{J}} \rightarrow [0, 1]$ such that $\Phi[(\mu_j)_{j \in \mathcal{J}}](\mathbf{A}) =$

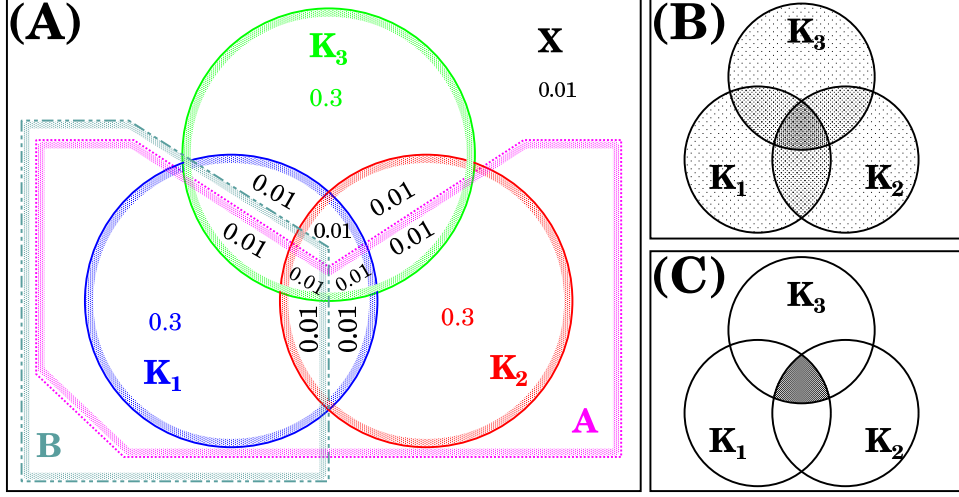


Figure 2: (A) A Venn diagram illustrating the *a priori* measure μ . Each number in the picture is the probability of the smallest region containing that number. (B) The LOP uses the average measure $\frac{1}{3}(\mu_1 + \mu_2 + \mu_3)$, where $\mu_j := \mu[\bullet | \mathbf{K}_j]$. (C) ‘Full disclosure’ uses the measure $\mu[\bullet | \mathbf{K}_{\mathcal{J}}]$, where $\mathbf{K}_{\mathcal{J}} := \mathbf{K}_1 \cap \mathbf{K}_2 \cap \mathbf{K}_3$.

j	$\mu[\mathbf{K}_j]$	$\mu[\mathbf{A} \cap \mathbf{K}_j]$	$\mu_j[\mathbf{A}]$	$\mu[(\mathbf{A}^c \cup \mathbf{B}) \cap \mathbf{K}_j]$	$\mu_j[\mathbf{A}^c \cup \mathbf{B}]$	$\mu[\mathbf{B} \cap \mathbf{K}_j]$	$\mu_j[\mathbf{B}]$
1	0.37	0.35	0.95	0.35	0.95	0.33	0.89
2	0.37	0.35	0.95	0.04	0.11	0.02	0.05
3	0.37	0.04	0.11	0.35	0.95	0.02	0.05
Linear Opinion Pool			0.67		0.67		0.33
Booleanization			T		T		F

Table 3: The conditional probabilities arising from Figure 2(A), the outcome of the LOP with $w_1 = w_2 = w_3 = 1/3$, and the result of ‘booleanizing’ the LOP with $0.33 < \theta_F \leq \theta_T < 0.67$. The ‘Discursive Dilemma’ rears its ugly head.

$F \left[(\mu_j[\mathbf{A}])_{j \in \mathcal{J}} \right]$ for all $\mathbf{A} \subseteq \mathbf{X}$. Then Φ must be an LOP (Wagner, 1982).

[ii] Suppose Φ is an SOPR such that

- (a) For any $\mathbf{A} \subseteq \mathbf{X}$, if $\mu_j[\mathbf{A}] = 0$ for all $j \in \mathcal{J}$, then $\Phi[(\mu_j)_{j \in \mathcal{J}}](\mathbf{A}) = 0$.
- (b) For any $\mathbf{A} \subseteq \mathbf{X}$, if $(\mu_j)_{j \in \mathcal{J}} \in \mathcal{P}(\mathbf{X})^{\mathcal{J}}$ and $(\nu_j)_{j \in \mathcal{J}} \in \mathcal{P}(\mathbf{X})^{\mathcal{J}}$ are such that $\mu_j[\mathbf{A}] = \nu_j[\mathbf{A}]$ for all $j \in \mathcal{J}$, then $\Phi[(\mu_j)_{j \in \mathcal{J}}](\mathbf{A}) = \Phi[(\nu_j)_{j \in \mathcal{J}}](\mathbf{A})$.

Then again Φ must be an LOP (McConway, 1981).

Note that the condition in [i] is the probabilistic analogue of the ‘Systematicity’ condition in List and Pettit (2002), while condition [ii](b) is the analogue of the ‘Independence’ condition in Dietrich and List (2007).

Unfortunately, the LOP has several deficiencies. For example, an SOPR Φ satisfies the *independence preservation property* (IPP) if, for any $\mathbf{A}, \mathbf{B} \subset \mathbf{X}$, if $\mu_j[\mathbf{A} \cap \mathbf{B}] = \mu_j[\mathbf{A}] \cdot \mu_j[\mathbf{B}]$ for all $j \in \mathcal{J}$, and $\mu_{\mathcal{J}} := \Phi[(\mu_j)_{j \in \mathcal{J}}]$, then $\mu_{\mathcal{J}}[\mathbf{A} \cap \mathbf{B}] = \mu_{\mathcal{J}}[\mathbf{A}] \cdot \mu_{\mathcal{J}}[\mathbf{B}]$. But the only LOP which satisfies the IPP is a *dictatorship* —i.e. there is some $j \in \mathcal{J}$ such that $w_j = 1$, while $w_i = 0$ for all $j \neq i$ (Lehrer and Wagner, 1983). When combined with the results [i] or [ii] above, this yields probabilistic versions of the impossibility theorems of List and Pettit (2002) and Dietrich and List (2007). For further shortcomings of the LOP, see Examples 2.1 and 2.2 below.

For these and other reasons, Baird (1985) and Loewer and Laddaga (1985) have rejected the LOP. However, to some extent the LOP's deficiencies reflect the inadequacies of SOPRs in general. For example if $\Phi : \mathcal{P}(\mathbf{X})^{\mathcal{J}} \rightarrow \mathcal{P}(\mathbf{X})$ is an SOPR defined by some function $F : [0, 1]^{\mathcal{J}} \rightarrow [0, 1]$ such that $\Phi[(\mu_j)_{j \in \mathcal{J}}](\mathbf{A}|\mathbf{B}) = F[(\mu_j[\mathbf{A}|\mathbf{B}])_{j \in \mathcal{J}}]$ for all $\mathbf{A}, \mathbf{B} \in \mathbf{X}$, then Φ must be a dictatorship (Dalkey, 1972, 1975). More generally, Genest and Wagner (1987) have shown that any SOPR which satisfies the IPP and has a rather general functional form (including any SOPR satisfying [i] above) must be a dictatorship; this yields yet another probabilistic version of List and Pettit's (2002) impossibility theorem. (Another problem appears in footnote #6 below).

In short: it is not possible to 'rationally' construct a collective probability distribution using *only* the data $(\mu_j)_{j \in \mathcal{J}}$. We need additional information about the private knowledge from which the jurors generate their subjective probability measures.

2 Full disclosure

If $\mathbf{K}_{\mathcal{J}} := \bigcap_{j \in \mathcal{J}} \mathbf{K}_j$, then we must have $x \in \mathbf{K}_{\mathcal{J}}$; in particular, this means that $\mathbf{K}_{\mathcal{J}} \neq \emptyset$. Thus, the best solution would be for the jurors to 'pool' their knowledge and define $\mu_{\mathcal{J}}[\mathbf{A}] := \mu[\mathbf{A}|\mathbf{K}_{\mathcal{J}}]$, $\mu_{\mathcal{J}}[\mathbf{B}] := \mu[\mathbf{B}|\mathbf{K}_{\mathcal{J}}]$, etc. Note that in general, this process cannot be described by an SOPR: there is no function $\Phi : \mathcal{P}(\mathbf{X})^{\mathcal{J}} \rightarrow \mathcal{P}(\mathbf{X})$ such that, for any family $\{\mathbf{K}_j\}_{j \in \mathcal{J}}$ of measurable subsets of \mathbf{X} with $\mu_j = \mu[\bullet|\mathbf{K}_j]$ for all $j \in \mathcal{J}$ and $\mathbf{K}_{\mathcal{J}} := \bigcap_{j \in \mathcal{J}} \mathbf{K}_j$, we will have $\Phi[(\mu_j)_{j \in \mathcal{J}}] = \mu[\bullet|\mathbf{K}_{\mathcal{J}}]$. [Compare Figures 2(B) and 2(C)]. In general, to determine the measure $\mu[\bullet|\mathbf{K}_{\mathcal{J}}]$, we must explicitly compute $\mathbf{K}_{\mathcal{J}}$; to do this, each juror j must fully disclose her private knowledge \mathbf{K}_j .

Example 2.1: Let $\mathbf{X} = [0, 1]^2$ (the unit box), let μ be the uniform measure. Let $\mathbf{W} := [0, \frac{1}{2}] \times [0, 1]$ and $\mathbf{E} := \mathbf{W}^c$ be the 'west' and 'east' halves of \mathbf{X} , while $\mathbf{S} := [0, 1] \times [0, \frac{1}{2}]$ and $\mathbf{N} := \mathbf{S}^c$ are the 'south' and 'north' halves of \mathbf{X} . Let $\mathbf{A} := \mathbf{S} \cap \mathbf{W}$ (the southwest quadrant), and let $\mathcal{J} := \{1, 2\}$. If $\mathbf{K}_1 = \mathbf{W}$ and $\mathbf{K}_2 = \mathbf{S}$, then $\mu_1[\mathbf{A}] = \mu[\mathbf{A}|\mathbf{W}] = 1/2$ and $\mu_2[\mathbf{A}] = \mu[\mathbf{A}|\mathbf{S}] = 1/2$. Thus, for any weights w_1, w_2 summing to 1, the LOP will estimate the probability of \mathbf{A} to be $w_1\mu_1[\mathbf{A}] + w_2\mu_2[\mathbf{A}] = (w_1 + w_2)(1/2) = 1/2$. But clearly, $\mu[\mathbf{A}|\mathbf{K}_1 \cap \mathbf{K}_2] = 1$; thus, if the jurors had disclosed their private knowledge, we would get a much better estimated probability for \mathbf{A} . \diamond

Sometimes it is impossible for the jurors to disclose their private knowledge. For example, we might be trying to aggregate the recorded opinions of experts who are widely separated in time and space. Even if the jurors are present in the same time and place, we can imagine that each \mathbf{K}_j represents some vast, poorly specified body of partly unconscious, intuitive and/or ineffable knowledge which would be impossible for each juror to disclose to the other jurors. Usually, the jurors can only disclose *some* of their knowledge. But sometimes, such ‘partial disclosure’ is sufficient.

Example 2.2: Let $\zeta : \mathbf{X} \rightarrow [0, 1]$ be a measurable function such that $\zeta(\mu) := \mu \circ \zeta^{-1}$ is the uniform measure on $[0, 1]$. For each $j \in \mathcal{J}$ and $m \in [1 \dots M_j]$, let $\zeta_m^j := \zeta + \epsilon_m^j$, where $\epsilon_m^j : \mathbf{X} \rightarrow \mathbb{R}$ are independent functions such that $\epsilon_m^j(\mu)$ is a normal distribution with mean 0 and variance $\sigma^2 \ll 1$. If $x_0 \in \mathbf{X}$ is the (unknown) state of nature, then the true value of $z := \zeta(x_0)$ is unknown (z is a random variable with uniform *a priori* distribution on $[0, 1]$). Suppose juror j knows $z_m^j := \zeta_m^j(x_0)$ for $m \in [1 \dots M_j]$ (i.e. these are her ‘measurements’ of z , with independent normal random measurement errors). Thus, $\mathcal{D}_j := \{z_m^j\}_{m=1}^{M_j}$ is j ’s ‘dataset’, and $\bar{z}^j := \frac{1}{M_j} \sum_{m=1}^{M_j} z_m^j$ is an ‘unbiased estimator’ for z , with standard error $\sigma/\sqrt{M_j}$.

Let $\mathbf{U} \subset [0, 1]$, and suppose the jury wants to estimate $\text{Prob}[z \in \mathbf{U}]$ using its measurement data. If $\mathbf{K}_j := \{x \in \mathbf{X} ; \zeta_m^j(x) = z_m^j, \forall m \in [1 \dots M_j]\}$ and $\mathbf{A} := \zeta^{-1}(\mathbf{U}) \subset \mathbf{X}$, then juror j estimates $\text{Prob}[z \in \mathbf{U} | \mathcal{D}_j] = \mu[\mathbf{A} | \mathbf{K}_j] \approx \nu_j[\mathbf{U}]$, where ν_j is the normal distribution with mean \bar{z}^j and variance σ^2/M_j .

The LOP would compute the average $\sum_{j=1}^J w_j \nu_j[\mathbf{U}]$ (for some weights $(w_j)_{j=1}^J$). A much better method is ‘full disclosure’: each juror to reveals her entire dataset \mathcal{D}_j , and the jury estimates $\text{Prob}[z \in \mathbf{U} | \bigcup_{j \in \mathcal{J}} \mathcal{D}_j] = \mu[\mathbf{A} | \bigcap_{j \in \mathcal{J}} \mathbf{K}_j] \approx \nu_{\mathcal{J}}[\mathbf{U}]$, where $\nu_{\mathcal{J}}$ is the normal distribution with mean $\bar{z}_{\mathcal{J}} := \frac{1}{M} \sum_{j \in \mathcal{J}} \sum_{m=1}^{M_j} z_m^j$ and variance σ^2/M , where $M := \sum_{j=1}^J M_j$. But ‘full disclosure’ is not required here. It suffices for each juror to disclose her estimator \bar{z}^j and her sample size M_j , because $M := \sum_{j=1}^J M_j$ and clearly $\bar{z}_{\mathcal{J}} = \frac{1}{M} \sum_{j \in \mathcal{J}} M_j \bar{z}^j$. Thus, given only $\{\bar{z}^j\}_{j=1}^J$ and $\{M_j\}_{j=1}^J$, the jury can determine $\nu_{\mathcal{J}}$ and then compute $\nu_{\mathcal{J}}[\mathbf{U}]$. \diamond

We will consider some other models of such ‘partial disclosure’ in §3 and §4.

3 Common Knowledge

Let $\mathbf{A} \subset \mathbf{X}$. Aumann (1976) showed that, if the probability estimate $\mu_j[\mathbf{A}]$ of each juror $j \in \mathcal{J}$ is the ‘common knowledge’ of all jurors, then all jurors must actually have *the same* probability estimate: there exists $\alpha \in [0, 1]$ such that $\mu_j[\mathbf{A}] = \alpha$ for all $j \in \mathcal{J}$. The obvious aggregate probability judgement is then $\mu_{\mathcal{J}}[\mathbf{A}] = \alpha$.

Formally, for all $j \in \mathcal{J}$, let \mathcal{K}_j be j ’s ‘knowledge partition’[¶], which we assume is *finite*. Let $x \in \mathbf{X}$ be the true worldstate, and suppose $x \in \mathbf{K}_j \in \mathcal{K}_j$. For any $\mathbf{A} \subset \mathbf{X}$, we say that

j *knows* \mathbf{A} in worldstate x if $\mathbf{K}_j \subseteq \mathbf{A}$. Let $\mathfrak{K}_j(\mathbf{A}) := \bigsqcup \{\mathbf{K} \in \mathcal{K}_j ; \mathbf{K} \subseteq \mathbf{A}\}$ be the event that j knows \mathbf{A} . Thus, $\mathfrak{K}_{\mathcal{J}}(\mathbf{A}) := \bigcap_{j \in \mathcal{J}} \mathfrak{K}_j(\mathbf{A})$ is the event that *everyone* knows \mathbf{A} (i.e. \mathbf{A} is ‘mutual knowledge’). Thus, $\mathfrak{K}_{\mathcal{J}}^2(\mathbf{A}) := \mathfrak{K}_{\mathcal{J}}[\mathfrak{K}_{\mathcal{J}}(\mathbf{A})]$ is the event that everyone *knows* that everyone knows \mathbf{A} , and so on. Finally $\mathfrak{K}_{\mathcal{J}}^\infty(\mathbf{A}) := \bigcap_{n=1}^\infty \mathfrak{K}_{\mathcal{J}}^n(\mathbf{A})$ is the event that \mathbf{A} is *common knowledge*.

For any $\alpha \in [0, 1]$ and $j \in \mathcal{J}$, let $\mathbf{M}_j(\mathbf{A}, \alpha) := \bigsqcup \{\mathbf{K} \in \mathcal{K}_j ; \mu[\mathbf{A}|\mathbf{K}_j] = \alpha\}$ be the event that j thinks the probability of \mathbf{A} is α . Aumann’s (1976) theorem states: for any $\alpha_1, \dots, \alpha_J \in [0, 1]$ and $x \in \mathbf{X}$, if the event $\mathbf{M}(\alpha_1, \dots, \alpha_J) := \mathbf{M}_1(\mathbf{A}, \alpha_1) \cap \dots \cap \mathbf{M}_J(\mathbf{A}, \alpha_J)$ is common knowledge at x , then $\alpha_1 = \dots = \alpha_J$. We call this an *Aumann consensus*. This suggests iteration of the following procedure, which we call ‘ $\Gamma\Pi$ ’:

- (Γ) Each juror publicly announces her current estimate of the probability of \mathbf{A} .
- (Π) Based on the stated estimates of the other jurors, each juror updates her own private estimate to account for this new ‘common knowledge’.

Geanakoplos and Polemarchakis (1982) showed that $\Gamma\Pi$ will iteratively converge to an Aumann consensus. This corroborates List and Pettit’s (2002, §4, p.101) suggestion that Discursive Dilemmas might be resolved through deliberation leading to ‘convergence’ of the juror’s beliefs.⁴

Example 3.1: Let $\mathbf{X} = [0, 1]^2$ with uniform measure μ , and suppose $\mathcal{K}_1 = \{\mathbf{N}, \mathbf{S}\}$ and $\mathcal{K}_2 = \{\mathbf{E}, \mathbf{W}\}$, where $\mathbf{N}, \mathbf{S}, \mathbf{E}, \mathbf{W} \subset \mathbf{X}$ are as in Example 2.1. Let $x \in \mathbf{S} \cap \mathbf{W}$ be the unknown true state of nature; hence $\mathbf{K}_1 = \mathbf{W}$ and $\mathbf{K}_2 = \mathbf{S}$.

If $\mathbf{A} = \mathbf{S} \cap \mathbf{W}$, then $\mu_1[\mathbf{A}] = \mu[\mathbf{A}|\mathbf{W}] = 1/2$ and $\mu_2[\mathbf{A}] = \mu[\mathbf{A}|\mathbf{S}] = 1/2$. Once juror 2 announces $\mu_2[\mathbf{A}] = 1/2$ [Step (Γ)], juror 1 will realize [Step (Π)] that $x \in \mathbf{S}$ (because otherwise, juror 2 would have said $\mu_2[\mathbf{A}] = 0$). Likewise, once juror 1 announces $\mu_1[\mathbf{A}] = 1/2$, juror 2 will realize that $x \in \mathbf{W}$. Thus, after publicly announcing their probability estimates during the first round of $\Gamma\Pi$, both jurors recognize that $x \in \mathbf{A}$; hence during round 2, both will agree that $\mu_{\mathcal{J}}[\mathbf{A}] = 1$. \diamond

Example 3.1 shows how $\Gamma\Pi$ works: after Step (Γ), each juror’s knowledge partition has effectively been *refined* to[¶] $\mathcal{K}_1 \vee \mathcal{K}_2 := \{\mathbf{S} \cap \mathbf{W}, \mathbf{S} \cap \mathbf{E}, \mathbf{N} \cap \mathbf{W}, \mathbf{N} \cap \mathbf{E}\}$, and both \mathbf{K}_1 and \mathbf{K}_2 have been updated to become $\mathbf{S} \cap \mathbf{W}$. Example 3.1 also shows that a naïve initial consensus (without common knowledge) is not the same thing as an Aumann consensus (*with* common knowledge). The juror’s probability estimates initially ‘agree’, but once they become common knowledge, both are revised upwards to a more accurate consensus. Geanakoplos and Polemarchakis (1982; Prop.4) showed that, generically, $\Gamma\Pi$ will converge in one step to the same Aumann consensus as the ideal of ‘full disclosure’ described in §2. But while this outcome is highly probable, it is not guaranteed (1982; Prop.3):

⁴Such ‘deliberative convergence’ is also discussed in List (2002, 2004, 2007). However, List (2008) has shown that restricting jurors to Boolean truth-valuations can seriously impede deliberative convergence; this contrasts sharply with the probabilistic convergence of ($\Gamma\Pi$).

Example 3.2: Continuing the notation of Example 3.1, suppose $\mathbf{B} := (\mathbf{S} \cap \mathbf{W}) \sqcup (\mathbf{N} \cap \mathbf{E})$; Then $\mu_1[\mathbf{B}] = 1/2 = \mu_2[\mathbf{B}]$. But after these estimates are disclosed [Step (Γ)], neither juror has any more information than before, so she will not revise her estimate during (Π). Thus, the resulting Aumann consensus will be $\mu_1[\mathbf{B}] = 1/2 = \mu_2[\mathbf{B}]$. However, if the jurors had fully disclosed their private knowledge, they would realize that $x \in \mathbf{K}_1 \cap \mathbf{K}_2 = \mathbf{S} \cap \mathbf{W}$, and the Aumann consensus would be $\mu[\mathbf{B}|\mathbf{S} \cap \mathbf{W}] = 1$. \diamond

If juror's knowledge partition \mathcal{K}_j is an uncountable sigma-algebra, then Aumann's result is still true (if we treat two events as 'equivalent' when they only differ by a null set), and the $\Gamma\Pi$ procedure still converges to consensus (via the martingale convergence theorem), in the limit as time $\rightarrow \infty$ (Nielsen, 1984, Thm. 4.1 & 4.2). Also, McKelvey and Page (1986; Thm.2) have shown that it is not necessary for the jurors to announce their probability estimates in Step (Γ); it suffices for there to be common knowledge of some (sufficiently informative) aggregate statistic of these estimates (e.g. their average) —even one generated from information inadvertently revealed by each juror's strategic behaviour —e.g. a market price. (For generalizations, see (Ménager, 2008) and the references therein.) Thus, to attain Aumann consensus, the jurors need not directly reveal any information to one another. Nor must any juror *trust* any other juror not to lie, or trust that every other juror trusts *her*, etc. (But they must still trust each other's competency, and they must be contemporaneous).

Finally, perfect knowledge is not necessary. For any $\mathbf{A} \subset \mathbf{X}$, $j \in \mathcal{J}$, and $p \in [0, 1]$, let $\mathfrak{B}_{p,j}(\mathbf{A}) := \bigsqcup \{\mathbf{K} \in \mathcal{K}_j ; \mu[\mathbf{A}|\mathbf{K}_j] \geq p\}$ be the event that j thinks the probability of \mathbf{A} is at least p (i.e. \mathbf{A} is a ' p -belief' of j). Thus $\mathfrak{B}_{p,\mathcal{J}}(\mathbf{A}) := \bigcap_{j \in \mathcal{J}} \mathfrak{B}_{p,j}(\mathbf{A})$ is the event that \mathbf{A} is a 'mutual p -belief'. For all $n \in \mathbb{N}$, let $\mathfrak{B}_{p,\mathcal{J}}^{n+1}(\mathbf{A}) := \mathfrak{B}_{p,\mathcal{J}}[\mathfrak{B}_{p,\mathcal{J}}^n(\mathbf{A})]$; then $\mathfrak{B}_{p,\mathcal{J}}^\infty(\mathbf{A}) := \bigcap_{n=1}^\infty \mathfrak{B}_{p,\mathcal{J}}^n(\mathbf{A})$ is the event that \mathbf{A} is *common p -belief*. Monderer and Samet (1989; Thm.A) proved: if $\mathbf{M}(\alpha_1, \dots, \alpha_J)$ is common p -belief, where $p = 1 - \epsilon$, then $|\alpha_i - \alpha_j| \leq 2\epsilon$ for all $i, j \in \mathcal{J}$.

The theory of common knowledge (or 'interactive epistemology') is now quite extensive; see (Geanakoplos, 1994) or (Lipman, 1999). But the predictions of Aumann consensus contradict empirical evidence: people who respect and trust each other often *continue* to disagree about the probability of certain events, even after ample dialogue. Perhaps the model's assumptions are unrealistic. For example, it assumes the jurors have a common prior probability measure (the so-called *Harsanyi Doctrine*); indeed, Feinberg (2000) has shown this is necessary for Aumann's result. But the Harsanyi Doctrine is not uncontroversial (Morris, 1995).

Or, perhaps Aumann consensus is possible in *theory*, but not in practice, because to attain it, each juror must mentally compute $\mathfrak{K}_{\mathcal{J}}^n[\mathbf{M}(\alpha_1, \dots, \alpha_J)]$ for arbitrarily large $n \in \mathbb{N}$; it is unclear whether ordinary people are willing (or able) to do this, so perhaps it is unsurprising that they do not reach a consensus, even if they theoretically could (Aumann, 1992). Also, to compute $\mathfrak{K}_{\mathcal{J}}^n[\mathbf{M}(\alpha_1, \dots, \alpha_J)]$, each juror needs perfect 'inter-subjective metaknowledge' of the knowledge partition \mathcal{K}_j of every other juror $j \in \mathcal{J}$, and also 'metametaknowledge' of other jurors' metaknowledge, etc. This so-called *Harsanyi-*

Aumann doctrine can be analyzed by embedding \mathbf{X} within a larger statespace \mathbf{X}^* ; each element of \mathbf{X}^* is a (transfinite) hierarchy representing the state of nature (in \mathbf{X}), each juror’s knowledge partition on \mathbf{X} , her metaknowledge partition (concerning other juror’s knowledge partitions on \mathbf{X}), her metametaknowledge partition (concerning other juror’s metaknowledge partitions), etc. In this formalism, a strict version of the Harsanyi-Aumann doctrine is false, but certain ‘approximate’ versions are true; see (Fagin et al., 1992, 1999) or (Geanakoplos, 1994, §15).

Example 3.1 shows how $\Gamma\Pi$ uses intersubjective metaknowledge: after learning $\mu_2[\mathbf{A}]$, juror 1 uses her metaknowledge of \mathcal{K}_2 to refine her own knowledge partition from \mathcal{K}_1 to $\mathcal{K}_1 \vee \mathcal{K}'_2$, for some $\mathcal{K}'_2 \preceq \mathcal{K}_2$. But juror 1 may have incomplete metaknowledge about 2 —perhaps only a coarser partition $\tilde{\mathcal{K}}_2 \prec \mathcal{K}_2$. Then no matter how much 1 learns about 2’s private knowledge via $\Gamma\Pi$, juror 1 can never refine her own knowledge partition beyond $\hat{\mathcal{K}}_1 := \mathcal{K}_1 \vee \tilde{\mathcal{K}}_2$. Conversely, if 2’s metaknowledge of 1 is $\tilde{\mathcal{K}}_1 \prec \mathcal{K}_1$, then 2 can never refine her own knowledge partition beyond $\hat{\mathcal{K}}_2 := \tilde{\mathcal{K}}_1 \vee \mathcal{K}_2$.

For \mathbf{A} to be common knowledge in world-state x , there must exist some $\mathbf{A}' \in \hat{\mathcal{K}}_1 \wedge \hat{\mathcal{K}}_2$ with $x \in \mathbf{A}' \subseteq \mathbf{A}$. But if $\tilde{\mathcal{K}}_j \neq \mathcal{K}_j$ (for $j = 1, 2$), then $\hat{\mathcal{K}}_1 \wedge \hat{\mathcal{K}}_2$ may be coarser than $\mathcal{K}_1 \vee \mathcal{K}_2$. If $\hat{\mathcal{K}}_1 \wedge \hat{\mathcal{K}}_2$ is a coarse partition, then only relatively large (i.e. low-information) subsets of \mathbf{X} can be common knowledge; hence people might agree on ‘obvious’ facts like whether it is day or night, but be unable to achieve Aumann consensus about the probability that CO_2 will exceed 710 ppm in 2100 AD.

4 Independent Confirmation

In Example 2.1, the LOP failed to recognize that \mathbf{K}_1 and \mathbf{K}_2 provide *independent* information about \mathbf{A} . The fact that both jurors *independently* confirm \mathbf{A} should raise the jury’s probability estimate for \mathbf{A} well above $\frac{1}{2}$. In contrast, suppose we knew that two jurors have very similar background knowledge —formally, suppose $\mathbf{K}_i \approx \mathbf{K}_j$. (For instance, in Example 2.2, suppose the datasets \mathcal{D}_i and \mathcal{D}_j were highly correlated —e.g. the errors ϵ_m^i and ϵ_m^j were not independent.) Then i and j ’s probability estimates for \mathbf{A} will be very close; any mechanism which treats i and j equally will incorrectly count this information ‘twice’. If $\mathbf{K}_i \approx \mathbf{K}_j$, then one of i or j is essentially redundant; a good methodology would (almost) entirely discount one of the two.

The LOP also pays insufficient heed to jurors with exceptionally high-quality information. Let $\mathbf{K}_{\mathcal{J}} := \bigcap_{j \in \mathcal{J}} \mathbf{K}_j$. If each juror’s private knowledge is correct, then we must have $x \in \mathbf{K}_{\mathcal{J}}$. Now suppose there is some $j \in \mathcal{J}$ with $\mu_j[\mathbf{A}] = 0$. Then $\mu[\mathbf{A} \cap \mathbf{K}_j] = 0$, so $\mu[\mathbf{A} \cap \mathbf{K}_{\mathcal{J}}] = 0$, so we should set $\mu_{\mathcal{J}}[\mathbf{A}] := 0$. However, if $\mu_i[\mathbf{A}] > 0$ for some $i \in \mathcal{J}$ with $w_i > 0$, then the LOP will incorrectly estimate $\mu_{\mathcal{J}}[\mathbf{A}] > 0$. Similarly, if $\mu_j[\mathbf{A}] = 1$, then we should set $\mu_{\mathcal{J}}[\mathbf{A}] := 1$; but if $\mu_i[\mathbf{A}] < 1$ for some $i \in \mathcal{J}$ with $w_i > 0$, then the LOP will incorrectly estimate $\mu_{\mathcal{J}}[\mathbf{A}] < 1$.

More generally, if there is some $j \in \mathcal{J}$ with $\mu_j[\mathbf{A}] \approx 1$ (or 0), then \mathbf{K}_j provides exceptionally high-quality information, which alone virtually guarantees the truth (or falsehood)

of proposition A ; the jury should thus give extra weight to j 's opinion. However, the LOP always gives j the same weight w_j , regardless of her information.

In Example 2.1, we *want* to compute $\mu[\mathbf{A}|\mathbf{K}_1 \cap \mathbf{K}_2]$. But suppose we cannot determine $\mathbf{K}_1 \cap \mathbf{K}_2$ (because \mathbf{K}_1 or \mathbf{K}_2 is hidden information) —all we know are the juror's probability estimates $\mu[\mathbf{A}|\mathbf{K}_1]$ and $\mu[\mathbf{A}|\mathbf{K}_2]$. We must estimate $\mu[\mathbf{A}|\mathbf{K}_1 \cap \mathbf{K}_2]$, given only knowledge of $\mu[\mathbf{A}|\mathbf{K}_1]$ and $\mu[\mathbf{A}|\mathbf{K}_2]$, and perhaps some information about the extent to which \mathbf{K}_1 and \mathbf{K}_2 provide 'independent' knowledge about \mathbf{A} , or the extent to which \mathbf{K}_1 provides 'better' information than \mathbf{K}_2 . For any $\mathbf{K}_1, \mathbf{K}_2 \subset \mathbf{X}$, we define the *correlation* between \mathbf{K}_1 and \mathbf{K}_2 by

$$C(\mathbf{K}_1, \mathbf{K}_2) := \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2]}{\mu[\mathbf{K}_1] \cdot \mu[\mathbf{K}_2]} = \frac{\mu[\mathbf{K}_1|\mathbf{K}_2]}{\mu[\mathbf{K}_1]} = \frac{\mu[\mathbf{K}_2|\mathbf{K}_1]}{\mu[\mathbf{K}_2]}.$$

Thus, $C(\mathbf{K}_1, \mathbf{K}_2) = 1$ iff \mathbf{K}_1 and \mathbf{K}_2 are independent[¶]. If $\mathbf{A} \subset \mathbf{X}$ is some other subset of \mathbf{X} , and $\mu[\mathbf{K}_1|\mathbf{A}]$ and $\mu[\mathbf{K}_2|\mathbf{A}]$ are both nonzero, then we define the *\mathbf{A} -conditional correlation* of \mathbf{K}_1 and \mathbf{K}_2 by

$$C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2) := \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2|\mathbf{A}]}{\mu[\mathbf{K}_1|\mathbf{A}] \cdot \mu[\mathbf{K}_2|\mathbf{A}]} = \frac{\mu[\mathbf{K}_i|\mathbf{K}_j \cap \mathbf{A}]}{\mu[\mathbf{K}_i|\mathbf{A}]}, \quad (i \neq j). \quad (1)$$

Thus, $C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2) = 1$ iff \mathbf{K}_1 and \mathbf{K}_2 are \mathbf{A} -conditionally independent. If $\mu[\mathbf{K}_1|\mathbf{A}] = 0$ or $\mu[\mathbf{K}_2|\mathbf{A}] = 0$, then the expression (1) is not well-defined; in this case, we define $C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2) := C(\mathbf{K}_1, \mathbf{K}_2)$. Thus, the ratio $C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2)/C(\mathbf{K}_1, \mathbf{K}_2)$ will be equal to one if either (a) One of \mathbf{K}_1 or \mathbf{K}_2 is disjoint from \mathbf{A} ; (b) \mathbf{K}_1 and \mathbf{K}_2 are independent, and *remain* independent when conditioned on \mathbf{A} ; or, (c) \mathbf{K}_1 and \mathbf{K}_2 are 'correlated' somehow, but the degree of correlation *does not change* when we condition on \mathbf{A} .

More generally, for any $\mathbf{K}_1, \dots, \mathbf{K}_J \subset \mathbf{X}$, we define

$$C(\mathbf{K}_1, \dots, \mathbf{K}_J) := \frac{\mu[\mathbf{K}_1 \cap \dots \cap \mathbf{K}_J]}{\mu[\mathbf{K}_1] \cdots \mu[\mathbf{K}_J]}; \quad \text{and}$$

$$C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J) := \begin{cases} \frac{\mu[\mathbf{K}_1 \cap \dots \cap \mathbf{K}_J|\mathbf{A}]}{\mu[\mathbf{K}_1|\mathbf{A}] \cdots \mu[\mathbf{K}_J|\mathbf{A}]}, & \text{if } \mu[\mathbf{K}_j|\mathbf{A}] > 0 \text{ for all } j \in \mathcal{J}; \\ C(\mathbf{K}_1, \dots, \mathbf{K}_J) & \text{if } \mu[\mathbf{K}_j|\mathbf{A}] = 0 \text{ for some } j \in \mathcal{J}. \end{cases}$$

Thus, $C(\mathbf{K}_1, \dots, \mathbf{K}_J) = 1$ if $\mathbf{K}_1, \dots, \mathbf{K}_J$ are jointly independent, and $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J) = 1$ if $\mathbf{K}_1, \dots, \mathbf{K}_J$ are \mathbf{A} -conditionally jointly independent.

Proposition 4.1 *Let $\mathbf{A}, \mathbf{K}_1, \dots, \mathbf{K}_J \subset \mathbf{X}$. Then*

$$\mu[\mathbf{A}|\mathbf{K}_1 \cap \dots \cap \mathbf{K}_J] = \frac{\mu[\mathbf{A}|\mathbf{K}_1] \cdots \mu[\mathbf{A}|\mathbf{K}_J]}{\mu[\mathbf{A}]^{J-1}} \cdot \frac{C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)}{C(\mathbf{K}_1, \dots, \mathbf{K}_J)}. \quad (2)$$

Thus, if $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J) = C(\mathbf{K}_1, \dots, \mathbf{K}_J)$, then

$$\mu[\mathbf{A}|\mathbf{K}_1 \cap \dots \cap \mathbf{K}_J] = \frac{\mu[\mathbf{A}|\mathbf{K}_1] \cdots \mu[\mathbf{A}|\mathbf{K}_J]}{\mu[\mathbf{A}]^{J-1}}. \quad (3)$$

Proof: If $\mu[\mathbf{K}_j|\mathbf{A}] > 0$ for all $j \in \mathcal{J}$, then this is a straightforward computation:

$$\begin{aligned}
& \frac{\mu[\mathbf{A}|\mathbf{K}_1] \cdots \mu[\mathbf{A}|\mathbf{K}_J]}{\mu[\mathbf{A}]^{J-1}} \cdot \frac{C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)}{C(\mathbf{K}_1, \dots, \mathbf{K}_J)} \\
&= \frac{\mu[\mathbf{A} \cap \mathbf{K}_1] \cdots \mu[\mathbf{A} \cap \mathbf{K}_J]}{\mu[\mathbf{A}]^{J-1} \cdot \mu[\mathbf{K}_1] \cdots \mu[\mathbf{K}_J]} \cdot \frac{\mu[\mathbf{K}_1] \cdots \mu[\mathbf{K}_J]}{\mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J]} \cdot \frac{\mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J|\mathbf{A}]}{\mu[\mathbf{K}_1|\mathbf{A}] \cdots \mu[\mathbf{K}_J|\mathbf{A}]} \\
&= \frac{\mu[\mathbf{A} \cap \mathbf{K}_1] \cdots \mu[\mathbf{A} \cap \mathbf{K}_J]}{\mu[\mathbf{A}]^{J-1} \cdot \mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J]} \cdot \frac{\mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J \cap \mathbf{A}] \cdot \mu[\mathbf{A}]^J}{\mu[\mathbf{K}_1 \cap \mathbf{A}] \cdots \mu[\mathbf{K}_J \cap \mathbf{A}] \cdot \mu[\mathbf{A}]} \\
&= \frac{\mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J \cap \mathbf{A}]}{\mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J]} = \mu[\mathbf{A}|\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J].
\end{aligned}$$

If $\mu[\mathbf{K}_j|\mathbf{A}] = 0$ for some $j \in \mathcal{J}$, then also $\mu[\mathbf{A}|\mathbf{K}_j] = 0$; thus $\mu[\mathbf{A}|\mathbf{K}_1] \cdots \mu[\mathbf{A}|\mathbf{K}_J] = 0$ and $\mu[\mathbf{A}|\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J] = 0$. Meanwhile, $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)/C(\mathbf{K}_1, \dots, \mathbf{K}_J) = 1$ because $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J) = C(\mathbf{K}_1, \dots, \mathbf{K}_J)$ by definition. Thus, equation (2) becomes “ $0 = 0/\mu[\mathbf{A}]^{J-1}$ ”, which is clearly true. \square

Example 4.2: (a) In Example 3.1, we have $J = 2$, and $C(\mathbf{K}_1, \mathbf{K}_2) = 1 = C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2)$. Thus, eqn.(2) becomes

$$\mu[\mathbf{A}|\mathbf{K}_1 \cap \mathbf{K}_2] = \frac{\mu[\mathbf{A}|\mathbf{K}_1] \cdot \mu[\mathbf{A}|\mathbf{K}_2]}{\mu[\mathbf{A}]} \cdot \frac{C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2)}{C(\mathbf{K}_1, \mathbf{K}_2)} = \frac{(1/2) \cdot (1/2)}{(1/4)} \cdot \frac{1}{1} = 1.$$

(b) In Example 3.2, $C(\mathbf{K}_1, \mathbf{K}_2) = 1$, but $C_{\mathbf{B}}(\mathbf{K}_1, \mathbf{K}_2) = 2$. Thus, eqn.(2) becomes

$$\mu[\mathbf{B}|\mathbf{K}_1 \cap \mathbf{K}_2] = \frac{\mu[\mathbf{B}|\mathbf{K}_1] \cdot \mu[\mathbf{B}|\mathbf{K}_2]}{\mu[\mathbf{B}]} \cdot \frac{C_{\mathbf{B}}(\mathbf{K}_1, \mathbf{K}_2)}{C(\mathbf{K}_1, \mathbf{K}_2)} = \frac{(1/2) \cdot (1/2)}{(1/2)} \cdot \frac{2}{1} = 1.$$

\diamond

However, if it's impossible for juror j to share her private knowledge \mathbf{K}_j , then we probably don't know enough about $\mathbf{K}_1, \dots, \mathbf{K}_J$ to compute $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)$ and $C(\mathbf{K}_1, \dots, \mathbf{K}_J)$. For all $j \in \mathcal{J}$, let \mathcal{K}_j be a finite or countable partition of \mathbf{X} , which we regard as j 's ‘knowledge partition’[¶]. If j 's knowledge is private, then we know that $\mathbf{K}_j \in \mathcal{K}_j$, but the actual value of \mathbf{K}_j is (for us) a \mathcal{K}_j -valued random variable; $\forall \mathbf{K} \in \mathcal{K}_j$, $\text{Prob}[\mathbf{K}_j = \mathbf{K}] = \mu[\mathbf{K}]$. Thus, $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)$ and $C(\mathbf{K}_1, \dots, \mathbf{K}_J)$ are also random variables. For any $\alpha \in [0, 1]$ and $j \in \mathcal{J}$, let $\mathcal{K}_j(\mathbf{A}, \alpha) := \{\mathbf{K} \in \mathcal{K}_j ; \mu[\mathbf{A}|\mathbf{K}] = \alpha\}$. Let $\mu_j[\mathbf{A}] := \mu[\mathbf{A}|\mathbf{K}_j]$. If $\mu_j[\mathbf{A}] = \alpha_j$ for some $\alpha_j \in [0, 1]$, then we know that $\mathbf{K}_j \in \mathcal{K}_j(\mathbf{A}, \alpha_j)$, even if we don't know what \mathbf{K}_j is. Thus, if we are told that $\mu_j[\mathbf{A}] = \alpha_j$ for all $j \in \mathcal{J}$ (for some $\{\alpha_j\}_{j \in \mathcal{J}} \subset [0, 1]$), then the expected value of the ratio $C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)/C(\mathbf{K}_1, \dots, \mathbf{K}_J)$, given this information, is

$$\begin{aligned}
EC_{\mathbf{A}}(\alpha_1, \dots, \alpha_J) &:= \frac{1}{M} \sum_{\mathbf{K}_1 \in \mathcal{K}_1(\mathbf{A}, \alpha_1)} \cdots \sum_{\mathbf{K}_J \in \mathcal{K}_J(\mathbf{A}, \alpha_J)} \mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J] \frac{C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)}{C(\mathbf{K}_1, \dots, \mathbf{K}_J)}. \\
\text{where } M &:= \sum_{\mathbf{K}_1 \in \mathcal{K}_1(\mathbf{A}, \alpha_1)} \cdots \sum_{\mathbf{K}_J \in \mathcal{K}_J(\mathbf{A}, \alpha_J)} \mu[\mathbf{K}_1 \cap \cdots \cap \mathbf{K}_J]
\end{aligned} \tag{4}$$

Proposition 4.1 then yields the following expectation for $\mu_{\mathcal{J}}[\mathbf{A}] := \mu[\mathbf{A}|\mathbf{K}_1 \cap \dots \cap \mathbf{K}_J]$:

$$\mathbb{E} \left(\mu_{\mathcal{J}}[\mathbf{A}] \mid \mu_j[\mathbf{A}] = \alpha_j, \forall j \in \mathcal{J} \right) = EC_{\mathbf{A}}(\alpha_1, \dots, \alpha_J) \cdot \frac{\alpha_1 \cdots \alpha_J}{\mu[\mathbf{A}]^{J-1}}. \quad (5)$$

In some cases, it may not be possible even to compute $EC_{\mathbf{A}}(\alpha_1, \dots, \alpha_J)$, since we don't even know the sets $\mathcal{K}_j(\mathbf{A}, \alpha_j)$ for all $j \in \mathcal{J}$. In this case, we can crudely approximate $EC_{\mathbf{A}}(\alpha_1, \dots, \alpha_J)$ with the constant

$$EC_{\mathbf{A}} := \sum_{\mathbf{K}_1 \in \mathcal{K}_1} \cdots \sum_{\mathbf{K}_J \in \mathcal{K}_J} \mu[\mathbf{K}_1 \cap \dots \cap \mathbf{K}_J] \frac{C_{\mathbf{A}}(\mathbf{K}_1, \dots, \mathbf{K}_J)}{C(\mathbf{K}_1, \dots, \mathbf{K}_J)}. \quad (6)$$

and then approximate (5) with

$$\mathbb{E} \left(\mu_{\mathcal{J}}[\mathbf{A}] \mid \mu_j[\mathbf{A}] = \alpha_j, \forall j \in \mathcal{J} \right) \approx EC_{\mathbf{A}} \cdot \frac{\alpha_1 \cdots \alpha_J}{\mu[\mathbf{A}]^{J-1}}. \quad (7)$$

For example, let $\mathcal{J} = \{1, 2\}$, and suppose \mathbf{A} is \mathcal{K}_1 -conditionally independent[¶] of \mathcal{K}_2 . Then clearly $\mu_{\mathcal{J}}[\mathbf{A}] = \mu_1[\mathbf{A}]$. Thus, the next result is no surprise; but it helps to illustrate the meaning of equation (5) and gauge the accuracy of approximation (7).

Corollary 4.3 *Suppose \mathbf{A} is \mathcal{K}_1 -conditionally independent of \mathcal{K}_2 .*

(a) *For any $\alpha_1, \alpha_2 \in [0, 1]$, we have $EC_{\mathbf{A}}(\alpha_1, \alpha_2) = \mu[\mathbf{A}]/\alpha_2$. Thus, expression (5) becomes $\mathbb{E} \left(\mu_{\mathcal{J}}[\mathbf{A}] \mid \mu_1[\mathbf{A}] = \alpha_1 \text{ \& } \mu_2[\mathbf{A}] = \alpha_2 \right) = \alpha_1$.*

(b) *$EC_{\mathbf{A}} = \mu[\mathbf{A}] \cdot \mathbb{E}(1/\mu_2[\mathbf{A}])$, where we regard $\mu_2[\mathbf{A}]$ as a random variable with $\text{Prob}[\mu_2[\mathbf{A}] = \alpha] = \mu[\mathcal{K}_2(\mathbf{A}, \alpha)]$ for all $\alpha \in [0, 1]$. Thus, for any $\alpha_1, \alpha_2 \in [0, 1]$, approximation (7) becomes $\mathbb{E} \left(\mu_{\mathcal{J}}[\mathbf{A}] \mid \mu_1[\mathbf{A}] = \alpha_1 \text{ \& } \mu_2[\mathbf{A}] = \alpha_2 \right) \approx \alpha_1 \alpha_2 \cdot \mathbb{E}(1/\mu_2[\mathbf{A}])$.*

Proof: If $\mathbf{A} \perp_{\mathcal{K}_1} \mathcal{K}_2$, then for any $\mathbf{K}_1 \in \mathcal{K}_1$ and $\mathbf{K}_2 \in \mathcal{K}_2$, we have $\mu[\mathbf{A}|\mathbf{K}_1 \cap \mathbf{K}_2] = \mu[\mathbf{A}|\mathbf{K}_1]$.

This means

$$\frac{\mu[\mathbf{A} \cap \mathbf{K}_1 \cap \mathbf{K}_2]}{\mu[\mathbf{K}_1 \cap \mathbf{K}_2]} = \frac{\mu[\mathbf{A} \cap \mathbf{K}_1]}{\mu[\mathbf{K}_1]}, \quad \text{hence} \quad \frac{\mu[\mathbf{A} \cap \mathbf{K}_1 \cap \mathbf{K}_2]}{\mu[\mathbf{A} \cap \mathbf{K}_1]} = \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2]}{\mu[\mathbf{K}_1]}. \quad (8)$$

$$\begin{aligned} \text{Thus, } C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2) &= \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2 | \mathbf{A}]}{\mu[\mathbf{K}_1 | \mathbf{A}] \cdot \mu[\mathbf{K}_2 | \mathbf{A}]} = \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2 \cap \mathbf{A}] \cdot \mu[\mathbf{A}]}{\mu[\mathbf{K}_1 \cap \mathbf{A}] \cdot \mu[\mathbf{K}_2 \cap \mathbf{A}]} \\ &\stackrel{(8)}{=} \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2] \cdot \mu[\mathbf{A}]}{\mu[\mathbf{K}_1] \cdot \mu[\mathbf{K}_2 \cap \mathbf{A}]} = \frac{\mu[\mathbf{K}_1 \cap \mathbf{K}_2]}{\mu[\mathbf{K}_1] \cdot \mu[\mathbf{K}_2]} \cdot \frac{\mu[\mathbf{K}_2] \cdot \mu[\mathbf{A}]}{\mu[\mathbf{K}_2 \cap \mathbf{A}]} \\ &= C(\mathbf{K}_1, \mathbf{K}_2) \cdot \frac{\mu[\mathbf{K}_2] \cdot \mu[\mathbf{A}]}{\mu[\mathbf{K}_2 \cap \mathbf{A}]} = C(\mathbf{K}_1, \mathbf{K}_2) \cdot \frac{\mu[\mathbf{A}]}{\mu[\mathbf{A}|\mathbf{K}_2]}. \end{aligned} \quad (9)$$

$$\begin{aligned}
\text{So } EC_{\mathbf{A}}(\alpha_1, \alpha_2) &\stackrel{(4)}{=} \frac{1}{M} \sum_{\substack{\mathbf{K}_1 \in \mathcal{K}_1(\mathbf{A}, \alpha_1) \\ \mathbf{K}_2 \in \mathcal{K}_2(\mathbf{A}, \alpha_2)}} \mu[\mathbf{K}_1 \cap \mathbf{K}_2] \frac{C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2)}{C(\mathbf{K}_1, \mathbf{K}_2)} \\
&\stackrel{(9)}{=} \frac{1}{M} \sum_{\substack{\mathbf{K}_1 \in \mathcal{K}_1(\mathbf{A}, \alpha_1) \\ \mathbf{K}_2 \in \mathcal{K}_2(\mathbf{A}, \alpha_2)}} \mu[\mathbf{K}_1 \cap \mathbf{K}_2] \frac{\mu[\mathbf{A}]}{\mu[\mathbf{A}|\mathbf{K}_2]} \\
&= \frac{1}{M} \sum_{\substack{\mathbf{K}_1 \in \mathcal{K}_1(\mathbf{A}, \alpha_1) \\ \mathbf{K}_2 \in \mathcal{K}_2(\mathbf{A}, \alpha_2)}} \mu[\mathbf{K}_1 \cap \mathbf{K}_2] \frac{\mu[\mathbf{A}]}{\alpha_2} \stackrel{(4)}{=} \frac{\mu[\mathbf{A}]}{\alpha_2}. \tag{10}
\end{aligned}$$

$$\begin{aligned}
\text{Also, } EC_{\mathbf{A}} &\stackrel{(6)}{=} \sum_{\substack{\mathbf{K}_1 \in \mathcal{K}_1 \\ \mathbf{K}_2 \in \mathcal{K}_2}} \mu[\mathbf{K}_1 \cap \mathbf{K}_2] \frac{C_{\mathbf{A}}(\mathbf{K}_1, \mathbf{K}_2)}{C(\mathbf{K}_1, \mathbf{K}_2)} \stackrel{(9)}{=} \sum_{\substack{\mathbf{K}_1 \in \mathcal{K}_1 \\ \mathbf{K}_2 \in \mathcal{K}_2}} \mu[\mathbf{K}_1 \cap \mathbf{K}_2] \frac{\mu[\mathbf{A}]}{\mu[\mathbf{A}|\mathbf{K}_2]} \\
&= \sum_{\mathbf{K}_2 \in \mathcal{K}_2} \mu[\mathbf{K}_2] \frac{\mu[\mathbf{A}]}{\mu[\mathbf{A}|\mathbf{K}_2]} = \mu[\mathbf{A}] \cdot \mathbb{E} \left(\frac{1}{\mu_2[\mathbf{A}]} \right). \tag{11}
\end{aligned}$$

This establishes the first assertions in (a) and (b). The second assertions follow by setting $J := 2$ and substituting (10) into (5) and (11) into (7). \square

Exact conditional independence is rare. If \mathbf{A} is only ‘almost’ \mathcal{K}_1 -conditionally independent of \mathcal{K}_2 (i.e. if \mathcal{K}_1 ‘mostly’ subsumes \mathcal{K}_2 -information about \mathbf{A}), then Corollary 4.3(a) and continuity imply that $\mathbb{E}(\mu_{\mathcal{J}}[\mathbf{A}] \mid \mu_1[\mathbf{A}] = \alpha_1 \ \& \ \mu_2[\mathbf{A}] = \alpha_2) \approx \alpha_1$. Comparing Corollary 4.3(b) to 4.3(a), we see that the approximation (7) is good as long as \mathbf{A} is ‘almost independent’ of \mathcal{K}_2 , so that the random variable $\mu_2[\mathbf{A}]$ is ‘almost constant’ (and hence, $\mathbb{E}(1/\mu_2[\mathbf{A}]) \approx 1/\alpha_2$).

5 Booleanization and its perils

Suppose the jury must be ‘decisive’, and produce not a probability measure, but a ‘verdict’ (i.e. a Boolean truth-valuation) concerning the propositions. The obvious procedure is to fix ‘thresholds’ $0 \leq \theta_F \leq \theta_T \leq 1$, and define a *booleanization* function $V : [0, 1] \longrightarrow \{T, ?, F\}$ by

$$V(p) := \begin{cases} F & \text{if } p < \theta_F; \\ ? & \text{if } \theta_F \leq p < \theta_T; \\ T & \text{if } p \geq \theta_T. \end{cases} \tag{12}$$

However, it is often impossible to obtain a logically consistent truth-valuation via (12), except by assigning the truth value ‘?’ to one or more propositions. For example, suppose that juror j believes that $\mu_j[\mathbf{A}^c \cap \mathbf{B}] = 0$, $\mu_j[\mathbf{A}] = \frac{2}{3}$ and $\mu_j[\mathbf{B}] = \frac{1}{3}$. Then $\mu[\mathbf{A}^c \cup \mathbf{B}] = \mu[\mathbf{A}^c] + \mu[\mathbf{B}] = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$. If we booleanize using any $\frac{1}{3} < \theta_F \leq \theta_T \leq \frac{2}{3}$, then we re-encounter the Discursive Dilemma, as in the bottom row of Table 3.

We can avoid this inconsistency if we make the booleanization thresholds more ‘indecisive’ (i.e make θ_F smaller and/or θ_T larger), but then (12) yields a useless truth-value of ‘?’

for one or more propositions. Furthermore, no matter how indecisive we make the thresholds, it is always possible to construct some system of logical propositions complicated enough to force a logically inconsistent booleanization.

Could we avoid these problems using a different booleanization instead of (12), or perhaps by applying *different* booleanizations to different propositions? The answer is ‘no’, as we shall now see. Let $\mathcal{T} := \{T, F\}$, and let \mathbf{X} be a set of possible ‘states of nature’. Given $\mathbf{A}_1, \dots, \mathbf{A}_N \subset \mathbf{X}$, a *decisive booleanization* for $\mathbf{A}_1, \dots, \mathbf{A}_N$ is an ordered n -tuple $\mathbf{V} = (V_1, \dots, V_N)$, where, for each $n \in [1\dots N]$, $V_n : [0, 1] \rightarrow \mathcal{T}$ is some function. If μ is a probability measure on \mathbf{X} , then $v_n := V_n(\mu[\mathbf{A}_n])$ is the ‘truth value’ assigned to event \mathbf{A}_n by booleanizing the probability $\mu[\mathbf{A}_n]$. Let $\mathbf{V}(\mu) := (v_1, \dots, v_N) \in \mathcal{T}^N$, and define

$$\mathbf{X}_{\mathbf{V}(\mu)} := \left\{ x \in \mathbf{X} ; \forall n \in [1\dots N], x \in \mathbf{A}_n \text{ if } v_n = T, \text{ and } x \in \mathbf{A}_n^c \text{ if } v_n = F \right\}.$$

For any $x \in \mathbf{X}_{\mathbf{V}(\mu)}$, the truth value of the proposition “ $x \in \mathbf{A}_n$ ” is equal to v_n for every $n \in [1\dots N]$ —in other words, x is a ‘semantic instantiation’ of the truth valuation $\mathbf{V}(\mu)$. Thus, $\mathbf{X}_{\mathbf{V}(\mu)} \neq \emptyset$ only if the truth valuation $\mathbf{V}(\mu)$ is logically consistent. We say \mathbf{V} *guarantees consistency* if $\mathbf{X}_{\mathbf{V}(\mu)} \neq \emptyset$ for any probability measure μ on \mathbf{X} .

We say \mathbf{V} *respects certainty* if $V_n(0) = F$ and $V_n(1) = T$ for all $n \in [1\dots N]$. A *good booleanizer* is some rule which assigns, to any set \mathbf{X} and any finite collection of subsets $\mathbf{A}_1, \dots, \mathbf{A}_N \subset \mathbf{X}$, a decisive booleanization $\mathbf{V} = (V_1, \dots, V_N)$ which respects certainty, yet guarantees consistency.

Proposition 5.1 *There is no good booleanizer.*

Proof: (by contradiction). Let $\mathcal{A} = \{a_1, \dots, a_N\}$ be a finite set of propositions which are *strongly connected* in the terminology of (Dietrich and List, 2007, §3). A ‘truth valuation’ on \mathcal{A} is thus an element of $\mathcal{T}^{\mathcal{A}}$. Let $\mathbf{X} := \{\mathbf{x} \in \mathcal{T}^{\mathcal{A}} ; \mathbf{x} \text{ is a logically consistent valuation on } \mathcal{A}\}$, and let \mathcal{B} be the power set of \mathbf{X} . For all $n \in [1\dots N]$, let $\mathbf{A}_n := \{\mathbf{x} \in \mathbf{X} ; x_n = T\}$ be the set of all valuations that say proposition a_n is ‘true’. Supposing there *was* a good booleanizer, let $\mathbf{V} := (V_1, \dots, V_N)$ be a decisive booleanization for the sets $\mathbf{A}_1, \dots, \mathbf{A}_N \subset \mathbf{X}$, which respects certainty, yet guarantees consistency.

If \mathcal{J} is any jury, we define a (Boolean) judgment aggregation function $\Phi : \mathbf{X}^{\mathcal{J}} \rightarrow \mathbf{X}$ as follows. For any profile $(\mathbf{x}_j)_{j \in \mathcal{J}} \in \mathbf{X}^{\mathcal{J}}$ of juror’s truth valuations, let $\mu : \mathcal{B} \rightarrow [0, 1]$ be the probability measure on \mathbf{X} where, for all $\mathbf{y} \in \mathbf{X}$, $\mu\{\mathbf{y}\} := \#\{j \in \mathcal{J} ; \mathbf{x}_j = \mathbf{y}\}$. Thus, $\mu[\mathbf{A}_n] = \#\{j \in \mathcal{J} ; \mathbf{x}_j \in \mathbf{A}_n\}$ is the number of jurors who ‘believe’ proposition a_n . Now let $\Phi[(\mathbf{x}_j)_{j \in \mathcal{J}}] := \mathbf{V}(\mu) = (v_1, \dots, v_N)$, where $v_n := V_n(\mu[\mathbf{A}_n])$ for every $n \in [1\dots N]$.

Claim 1: $\mathbf{V}(\mu) \in \mathbf{X}$.

Proof: $\mathbf{V}(\mu)$ is the *unique* point in $\mathcal{T}^{\mathcal{A}}$ such that, for all $n \in [1\dots N]$, $\mathbf{V}(\mu) \in \mathbf{A}_n$ if $v_n = T$, while $\mathbf{V}(\mu) \in \mathbf{A}_n^c$ if $v_n = F$. Thus, $\mathbf{V}(\mu)$ is the only element of $\mathcal{T}^{\mathcal{A}}$ which *could* be in $\mathbf{X}_{\mathbf{V}(\mu)}$. But $\mathbf{X}_{\mathbf{V}(\mu)} \neq \emptyset$, because \mathbf{V} guarantees consistency —hence $\mathbf{X}_{\mathbf{V}(\mu)} = \{\mathbf{V}(\mu)\}$, which means that $\mathbf{V}(\mu) \in \mathbf{X}$. ◇ Claim 1

Claim 1 makes $\Phi : \mathbf{X}^{\mathcal{J}} \longrightarrow \mathbf{X}$ a well-defined function; in the terminology of Dietrich and List (2007), Φ satisfies *universal domain* (it is defined everywhere on $\mathbf{X}^{\mathcal{J}}$) and *collective rationality* (it maps into \mathbf{X}). Furthermore, Φ is *independent*, because the value of $v_n = V_n(\mu[\mathbf{A}_n])$ depends only on the number of jurors who believe proposition a_n (that is, $\mu[\mathbf{A}_n]$). This also means that Φ is *anonymous*: it treats all jurors the same. Finally, Φ *respects unanimity* because \mathbf{V} respects certainty. Thus, Theorem 2 of Dietrich and List (2007) says Φ must be a dictatorship —but this is impossible if $\#(\mathcal{J}) \geq 2$, because Φ is anonymous. Thus, we have a contradiction. \square

So, booleanization is generally impossible. Fortunately, it usually isn't even necessary. When a jury must be 'decisive', it normally must decide a *course of action* —it doesn't need to declare a Boolean verdict on every proposition which may be germane to this course of action. In an uncertain situation, the Bayesian decision procedure is to select the action which yields the highest *expected utility*. To perform this procedure, a rational jury must specify: (1) A probability distribution μ over the set \mathbf{X} of 'states of nature', and (2) A utility function $U : \mathcal{A} \times \mathbf{X} \longrightarrow \mathbb{R}$, where \mathcal{A} is the set of possible actions. Finding the $a^* \in \mathcal{A}$ which maximizes $\int_{\mathbf{X}} U(a^*, x) d\mu[x]$ is then merely a computational problem (difficult in practice, but solvable in principle). The real problem is for the jury to construct U and μ . The construction of U (the 'social welfare function') is the concern of traditional social choice theory.⁵ The construction of μ is our present concern.⁶ At no time during this process need the jury declare a Boolean verdict for any propositions corresponding to subsets of \mathbf{X} .

6 Consequentialism versus deontology

The 'Bayesian decision procedure' in §5 is 'consequentialist': it says that the best action is the one which yields the best consequences (e.g. the highest expected utility).⁷ However, 'deontological' philosophers argue that actions should be judged not by their consequences, but by their conformity with certain inviolable moral axioms, such as respect for inalienable

⁵The 'correct' social welfare function is a subject of ongoing controversy. However, Harsanyi (1955; 1977 §4.8) has shown that, if both the individual jurors and the jury as a whole are expected-utility maximizers, as we are suggesting here, and if the jury chooses Pareto-preferred alternatives whenever possible, then the jury's utility function must be a linear combination of the utility functions of the jurors. Mongin (1994) and others have extended Harsanyi's argument to derive a utilitarian social welfare function.

⁶Hylland and Zeckhauser (1979) have shown that, if we aggregate the jurors' utilities $\{u_j\}_{j \in \mathcal{J}}$ with a social welfare function, and also aggregate their probability measures $\{\mu_j\}_{j \in \mathcal{J}}$ with a nondictatorial SOPR, then the collective Bayesian choice a^* may be 'Pareto-inefficient': there may be some other $b \in \mathcal{A}$ such that $\int_{\mathbf{X}} u_j(b, x) d\mu_j[x] > \int_{\mathbf{X}} u_j(a^*, x) d\mu_j[x]$, $\forall j \in \mathcal{J}$. This can be seen as another argument against SOPRs.

⁷Notwithstanding footnote #5, 'consequentialist' does not mean 'welfarist' —the 'consequences' in question could refer to non-welfare goods like 'liberty' or 'autonomy'. Also, even 'welfarism' recognizes the 'instrumental' value of such non-welfare goods, because of their strong impact on welfare itself —see footnotes #10 and #11.

‘rights’ (e.g. property rights, personal liberties) and obedience to ‘duties’ (e.g. the law; the terms of a valid contract; the holy scriptures). For example, the juridical reasoning considered by Kornhauser and Sager is deontological. Deontological reasoning is a kind of deductive logic; hence all deontological propositions must be assigned Boolean truth values. Thus, a deontologist would argue that, when deciding a course of action, the jury in fact *does* need to declare a Boolean verdict on every proposition which may be germane to this course of action.

A full discussion of consequentialism versus deontology is outside the scope of this paper. We will simply note that, because it works with discrete categories (e.g. forbidden/allowed/obligatory, true/false) rather than continuous quantities (e.g. utility, probability), deontology already has two well-known defects:

- Deontology is incapable of making the (often tragic) tradeoffs which are ubiquitous in politics. (Example: If we are forced to choose, is it worse to violate the property rights of ten people, or the personal liberties of five?).⁸
- Deontology is incapable of dealing with risk, which is inevitable in any complex decision. (Example: is it worse to risk a 20% chance of trampling the liberties of ten people, or a 10% chance of trampling the liberties of twenty?).⁹

To these two defects, we can now add a third:

- By insisting on Boolean truth valuations, deontology inexorably collides with the Discursive Dilemma.

Deontological arguments still have a place in a consequentialist moral reasoning, but only as a kind of heuristic or shorthand:

1. They can help to make approximate, ‘order-of-magnitude’ estimates of the expected utility of actions, when precise computations are impossible (e.g. due to incomplete data) or excessively complicated (e.g. due to long-term consequences).¹⁰

⁸Each tradeoff can be resolved by adding some supplementary moral axiom (e.g. “the liberty of N people takes precedence over the property rights of $2N$ or less people”). But the accretion of such *ad hoc* axioms will inevitably lead to logical inconsistencies, eventually.

⁹Indeed, it has been known since von Neumann and Morgenstern (1947 [2007]) that, if an actor i makes risky choices which satisfy minimal conditions of ‘rationality’, then there is a utility function u_i such that i always acts to maximize the expected value of u_i . So, even a supposedly ‘deontological’ actor, if she responds ‘rationally’ to risk, is behaving *as if* she is a consequentialist, whether she knows it or not.

¹⁰Example: violating personal liberties usually causes massive disutilities. Lacking more precise data, we could approximate the ‘expected utility’ of any action which violated liberties as being ‘ $-\infty$ ’, compared to an action which did not; this has roughly the same effect as treating personal liberty as a ‘right’.

2. They can efficiently and concisely summarize complex consequentialist arguments (or at least, approximations of them).¹¹

However, it must be remembered that deontological reasoning is never more than a shortcut for a more nuanced consequentialist argument. When a deontological argument runs into trouble (e.g. tradeoffs, risk, or discursive dilemmas), it should be jettisoned in favour of a consequentialist (e.g. Bayesian) decision procedure.

Conclusion

We can mostly obviate the ‘Discursive Dilemma’ if we reconceive the problem of ‘judgment aggregation’ in three ways:

- (i) Aggregate the juror’s probability estimates, rather than Boolean truth valuations.
- (ii) Utilize the juror’s private information (from which they derive their probability estimates) in addition to the estimates themselves.
- (iii) Do not try to ‘booleanize’ the jury’s aggregated probability judgement (§5).

Section 1 showed that (i) alone is not sufficient; SOPRs (which ignore the juror’s private information) encounter ‘impossibility theorems’ very similar to the Discursive Dilemma. Sections 2-4 explored three models of (ii), from an unrealistic ideal of ‘full disclosure’ (§2), to a (perhaps equally unrealistic) model of ‘Aumann consensus’ arising from perfect intersubjective metaknowledge (§3), to a less ambitious mechanism which uses only a crude measure of the jurors’ ‘independence’ from one another (§4).

All three models show the importance of *deliberation*. It is simply not sufficient to mechanically aggregate judgements (probabilistic, Boolean, or otherwise) using some simple ‘voting rule’. It is necessary for the jurors to deliberate, to share information, and to revise their private judgements in light of the information revealed by others. Ideally, at the end of this deliberation, there will be no *need* to ‘aggregate’ judgements, because there will be unanimous consensus (§3). Even if there is still dissensus, we can use the additional information revealed by the jurors to refine the aggregation process (§4). Thus, our analysis provides strong support for the ‘deliberative’ conception of democracy which has been advanced by Dryzek and List (2003) and others.¹²

¹¹ Example: people must expect that their personal property will be protected, their contracts will be fulfilled, and they will be free to engage in economic transactions —otherwise the modern market economy would disintegrate, greatly diminishing prosperity and causing long-term disutility for everyone. To sustain these expectations, the State must always be seen to vigorously enforce contracts, protect personal property, and respect economic freedoms. In other words, the State must act *as if* we have a ‘moral duty’ to honour contracts, and a ‘right’ to property or economic liberty.

¹²See Bohman and Rehg (1997), Elster (1998), Dryzek (2002), Fishkin and Laslett (2003), Amsler (2004), Bächtiger and Steiner (2005), van Aaken et al. (2004) or Rosenberg (2007).

Appendix on Probability Theory

Let \mathbf{X} be a set of ‘world-states’. A *sigma-algebra* on \mathbf{X} is a collection \mathcal{B} of subsets of \mathbf{X} which is closed under complements, countable intersection, and countable unions. A subset $\mathbf{A} \subset \mathbf{X}$ is an *event* (or an *\mathcal{B} -measurable set*) if $\mathbf{A} \in \mathcal{B}$. A *probability measure* is a function $\mu : \mathcal{B} \rightarrow [0, 1]$, with $\mu[\mathbf{X}] = 1$, which is *countably additive* —i.e. $\mu[\bigsqcup_{n=1}^{\infty} \mathbf{A}_n] = \sum_{n=1}^{\infty} \mu[\mathbf{A}_n]$ for any disjoint events $\mathbf{A}_1, \mathbf{A}_2, \dots$, in \mathbf{X} . The ordered triple $(\mathbf{X}, \mathcal{B}, \mu)$ is then called a *probability space*. If $\mathbf{A}, \mathbf{K} \subset \mathbf{X}$, then $\mu[\mathbf{A}|\mathbf{K}] := \mu[\mathbf{A} \cap \mathbf{K}] / \mu[\mathbf{K}]$ is the *conditional probability* of \mathbf{A} , given \mathbf{K} .

A *partition* of \mathbf{X} is a collection $\mathcal{K} = \{\mathbf{K}_n\}_{n=1}^N$ of disjoint events such that $\mathbf{X} = \bigsqcup_{n=1}^N \mathbf{K}_n$ (where $N \in \mathbb{N} \sqcup \{\infty\}$). Typically, \mathcal{K} is the *knowledge partition* of some juror j : at any time, j ’s ‘knowledge set’ \mathbf{K}_j is some element of \mathcal{K} . For example, suppose j obtains her knowledge by conducting some ‘measurement’ or ‘experiment’, described by some measurable function $\phi_j : \mathbf{X} \rightarrow \mathbb{N}$. Then j ’s ‘knowledge’ of the unknown world-state $x_0 \in \mathbf{X}$ is the measurement value $N_j := \phi_j(x_0) \in \mathbb{N}$; hence $\mathbf{K}_j = \phi_j^{-1}\{N_j\}$. Thus, $\mathcal{K} = \{\phi_j^{-1}\{n\} ; n \in \mathbb{N}\}$.

If $\mathbf{A} \subset \mathbf{X}$, then \mathbf{A} is *\mathcal{K} -measurable* if $\mathbf{A} = \bigsqcup_{\mathbf{K} \in \mathcal{K}'} \mathbf{K}$ for some subset $\mathcal{K}' \subseteq \mathcal{K}$. This means that juror j ’s private knowledge alone provides complete information about \mathbf{A} : for any $\mathbf{K} \in \mathcal{K}$, either $\mu[\mathbf{A}|\mathbf{K}] = 1$ or $\mu[\mathbf{A}|\mathbf{K}] = 0$ (because either $\mathbf{K} \subseteq \mathbf{A}$, or \mathbf{K} is disjoint from \mathbf{A}). Let $\langle \mathcal{K} \rangle$ be the family of all \mathcal{K} -measurable subsets of \mathbf{X} (that is: $\langle \mathcal{K} \rangle$ is the sigma-algebra generated by \mathcal{K}).

Let \mathcal{K}_1 and \mathcal{K}_2 be the knowledge partitions of jurors 1 and 2. Say \mathcal{K}_1 *refines* \mathcal{K}_2 if $\mathcal{K}_2 \subseteq \langle \mathcal{K}_1 \rangle$. (Notation: “ $\mathcal{K}_2 \preceq \mathcal{K}_1$ ”). This means that juror 1’s knowledge completely subsumes juror 2’s knowledge. We define $\mathcal{K}_1 \wedge \mathcal{K}_2$ to be the coarsest partition which refines both \mathcal{K}_1 and \mathcal{K}_2 , and define $\mathcal{K}_1 \vee \mathcal{K}_2$ to be the finest partition which is refined by both \mathcal{K}_1 and \mathcal{K}_2 . If $\mathbf{A}, \mathbf{K} \subset \mathbf{X}$, then \mathbf{A} is *independent* of \mathbf{K} if $\mu[\mathbf{A}|\mathbf{K}] = \mu[\mathbf{A}]$. (Notation: “ $\mathbf{A} \perp \mathbf{K}$ ”). Likewise, \mathbf{A} is *independent* of \mathcal{K}_2 if $\mathbf{A} \perp \mathbf{K}_2$ for all $\mathbf{K}_2 \in \mathcal{K}_2$. (Notation: “ $\mathbf{A} \perp \mathcal{K}_2$ ”). This means that juror 2 is totally ignorant of \mathbf{A} —her private knowledge tells us absolutely nothing about \mathbf{A} . We say \mathcal{K}_1 is *independent* of \mathcal{K}_2 if $\mathbf{K}_1 \perp \mathcal{K}_2$ for every $\mathbf{K}_1 \in \mathcal{K}_1$. (Notation: “ $\mathcal{K}_1 \perp \mathcal{K}_2$ ”). This means that jurors 1 and 2 have completely disjoint information about the world.

Finally, we say that \mathbf{A} is *\mathcal{K}_1 -conditionally independent* of \mathcal{K}_2 if, for any $\mathbf{K}_1 \in \mathcal{K}_1$ and $\mathbf{K}_2 \in \mathcal{K}_2$, $\mu[\mathbf{A}|\mathbf{K}_1 \cap \mathbf{K}_2] = \mu[\mathbf{A}|\mathbf{K}_1]$. (Notation: “ $\mathbf{A} \perp_{\mathcal{K}_1} \mathcal{K}_2$ ”). This means that juror 1’s knowledge of \mathbf{A} totally subsumes juror 2’s knowledge of \mathbf{A} : once we have juror 1’s opinion, juror 2’s opinion is redundant. For example: (a) If $\mathcal{K}_2 \preceq \mathcal{K}_1$, then $\mathbf{A} \perp_{\mathcal{K}_1} \mathcal{K}_2$. (b) If $\mathcal{K}_1 \perp \mathcal{K}_2$ and also $\mathbf{A} \perp \mathcal{K}_2$ (e.g. if $\mathbf{A} \in \langle \mathcal{K}_1 \rangle$), then $\mathbf{A} \perp_{\mathcal{K}_1} \mathcal{K}_2$. (c) Suppose $\mathcal{K}_1, \mathcal{K}_2$, and \mathbf{A} represent information about a Markov process. If \mathcal{K}_1 represents information about the process at time t , and \mathbf{A} represents an event which occurs *after* time t , while \mathcal{K}_2 represents information from *before* time t , then $\mathbf{A} \perp_{\mathcal{K}_1} \mathcal{K}_2$.

References

- Amsler, T. (Ed.), Winter 2004. Special issue on Deliberative Democracy. Vol. 93 #4 of National Civic Review. Wiley.
- Aumann, R. J., 1976. Agreeing to disagree. Ann. Statist. 4 (6), 1236–1239.
- Aumann, R. J., 1992. Irrationality in game theory. In: Dasgupta, P., Gale, D., Hart, O., Maskin, E. (Eds.), Economic analysis of markets and games. MIT Press, pp. 214–227.

- Bächtiger, A., Steiner, J. (Eds.), July and September 2005. Empirical Approaches to Deliberative Democracy. Vol. 40 #2 and #3 of Acta Politica. Palgrave MacMillan.
- Baird, D., 1985. Lehrer-wagner consensual probabilities do not adequately summarize the available information. *Synthese* 62, 47–62.
- Bohman, J., Rehg, W. (Eds.), 1997. Deliberative Democracy : Essays on Reason and Politics. MIT Press.
- Cholvy, L., 1998. Reasoning about merged information. In: Belief change. Vol. 3 of Handb. Defeasible Reason. Uncertain. Manag. Syst. Kluwer Acad. Publ., Dordrecht, pp. 233–263.
- Dalkey, N. C., 1972. An impossibility theorem for group probability functions. Tech. Rep. P-4862, The Rand Corporation, Santa Monica, CA.
- Dalkey, N. C., 1975. Toward a theory of group estimation. In: Linstone, H., , Turoff, M. (Eds.), The Delphi Method: Techniques and Applications. Addison-Wesley, Reading, MA.
- de Finetti, B., 1954. Media di decisioni e media di opinioni. *Bull. Inst. Internat. Statist.* 24 (2), 144–157.
- Dietrich, F., 2007. A generalised model of judgment aggregation. *Soc. Choice Welf.* 28 (4), 529–565.
- Dietrich, F., List, C., 2007. Arrow’s theorem in judgment aggregation. *Soc. Choice Welf.* 29 (1), 19–33.
- Dryzek, J., List, C., 2003. Social choice theory and deliberative democracy: a reconciliation. *British Journal of Political Science* 33 (1), 1–28.
- Dryzek, J. S., 2002. Deliberative Democracy and Beyond: Liberals, Critics, Contestations. Oxford UP, Oxford, UK.
- Elster, J. (Ed.), 1998. Deliberative Democracy. Cambridge UP.
- Fagin, R., Geanakoplos, J., Halpern, J. Y., Vardi, M. Y., 1992. The expressive power of the hierarchical approach to modeling knowledge and common knowledge. In: Vardi, M. Y. (Ed.), Fourth Symposium on theoretical aspects of reasoning about knowledge. Morgan Kaufmann, Los Altos, pp. 229–244.
- Fagin, R., Geanakoplos, J., Halpern, J. Y., Vardi, M. Y., 1999. The hierarchical approach to modeling knowledge and common knowledge. *Internat. J. Game Theory* 28 (3), 331–365.
- Feinberg, Y., 2000. Characterizing common priors in the form of posteriors. *J. Econom. Theory* 91 (2), 127–179.
- Fishkin, J. S., Laslett, P. (Eds.), 2003. Debating Deliberative Democracy. Blackwell.
- Gärdenfors, P., 2006. A representation theorem for voting with logical consequences. *Economics and Philosophy* 22, 181–190.
- Geanakoplos, J., 1994. Common knowledge. In: Handbook of game theory with economic applications, Vol. II. North-Holland, Amsterdam, pp. 1437–1496.
- Geanakoplos, J. D., Polemarchakis, H. M., 1982. We can’t disagree forever. *Journal of Economic Theory*, 28, 192–200.
- Genest, C., Wagner, C. G., 1987. Further evidence against independence preservation in expert judgement synthesis. *Aequationes Math.* 32 (1), 74–86.

- Genest, C., Zidek, J. V., 1986. Combining probability distributions: a critique and an annotated bibliography. *Statist. Sci.* 1 (1), 114–148.
- Harsanyi, J., 1955. Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. *Journal of Political Economy* 63, 309–321.
- Harsanyi, J., 1977. Rational behaviour and bargaining equilibrium in games and social situations. Cambridge UP, Cambridge, UK.
- Hylland, A., Zeckhauser, R., 1979. The impossibility of Bayesian group decision making with separate aggregation of beliefs and values. *Econometrica* 47 (6), 1321–1336.
- Konieczny, S., Pino Pérez, R., 2005. Propositional belief base merging or how to merge beliefs/goals coming from several sources and some links with social choice theory. *European J. Oper. Res.* 160 (3), 785–802.
- Kornhauser, L., Sager, L., 1986. Unpacking the court. *Yale Law Journal*.
- Kornhauser, L., Sager, L., 1993. The one and the many: adjudication in collegial courts. *California Law Review* 91, 1–51.
- Kornhauser, L., Sager, L., 2004. The many as one: Integrity and group choice in paradoxical cases. *Philosophy & Public Affairs* 32, 249–276.
- Lehrer, K., Wagner, C., 1981. Rational consensus in science and society: A philosophical and mathematical study. D. Reidel Publishing Co., Dordrecht.
- Lehrer, K., Wagner, C., 1983. Probability amalgamation and the independence issue: a reply to Laddaga. *Synthese* 55 (3), 339–346.
- Lipman, B. L. (Ed.), August 1999. Special Issue on Interactive Epistemology. Vol. 28 (3) of *International Journal of Game Theory*. Springer-Verlag, Heidelberg, pp.261–442.
- List, C., 2002. Two concepts of agreement. *The Good Society* 11 (1), 72–79; reprinted as List (2004).
- List, C., 2004. Substantive and meta-agreement. In: van Aaken et al. (2004), pp. 119–134.
- List, C., 2007. Deliberation and agreement. In: Rosenberg (2007).
- List, C., 2008. Group deliberation and the revision of judgments: An impossibility result. (preprint).
- List, C., Pettit, P., 2002. Aggregating sets of judgements: an impossibility result. *Economics and Philosophy* 18, 89–110.
- Loewer, B., Laddaga, R., 1985. Destroying the consensus. *Synthese* 62, 79–95.
- McConway, K. J., 1981. Marginalization and linear opinion pools. *J. Amer. Statist. Assoc.* 76 (374), 410–414.
- McKelvey, R. D., Page, T., 1986. Common knowledge, consensus, and aggregate information. *Econometrica* 54 (1), 109–127.
- Ménager, L., 2008. Consensus and common knowledge of an aggregate of decisions. *Games and Economic Behavior* 62, 722–731.

- Monderer, D., Samet, D., 1989. Approximating common knowledge with common beliefs. *Games Econom. Behav.* 1 (2), 170–190.
- Mongin, P., 1994. Harsanyi’s aggregation theorem: multi-profile version and unsettled questions. *Soc. Choice Welf.* 11 (4), 331–354.
- Morris, S., October 1995. The common prior assumption in economic theory. *Economics and Philosophy* 11 (2), 227–53.
- Nielsen, L. T., 1984. Common knowledge, communication, and convergence of beliefs. *Math. Social Sci.* 8 (1), 1–14.
- Pauly, M., van Hees, M., 2006. Logical constraints on judgement aggregation. *J. Philos. Logic* 35 (6), 569–585.
- Pigozzi, G., 2006. Belief merging and the discursive dilemma: an argument-based account to paradoxes of judgment aggregation. *Synthese* 152 (2), 285–298.
- Rosenberg, S. W. (Ed.), 2007. *Deliberation, Participation and Democracy: Can the People Govern?* Palgrave Macmillan.
- Savage, L. J., 1954. *The foundations of statistics*. John Wiley & Sons Inc., New York.
- Stone, M., 1961. The opinion pool. *Ann. Math. Statist.* 32, 1339–1342.
- van Aaken, A., List, C., Luetge, C. (Eds.), 2004. *Deliberation and Decision: Economics, constitutional theory and deliberative democracy*. Ashgate, Aldershot, UK.
- van Hees, M., 2007. The limits of epistemic democracy. *Soc. Choice Welf.* 28 (4), 649–666.
- von Neumann, J., Morgenstern, O., 1947 [2007]. *Theory of games and economic behavior*. Princeton University Press, Princeton, NJ.
- Wagner, C., 1982. Allocation, Lehrer models, and the consensus of probabilities. *Theory and Decision* 14 (2), 207–220.